DOCUMENT RESUME

ED 397 102 TM 025 224

AUTHOR Smith, Catherine
TITLE What Have We Learned from Assessing Employability Skills Portfolios?
PUB DATE 10 Apr 96
NOTE 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Employment Potential; Grade 11; High Schools; *High School Students; *Job Skills; *Portfolio Assessment; Portfolios (Background Materials); Professional Development; School Culture; School Districts; *Scoring; Student Evaluation; Testing Programs; Test Results
IDENTIFIERS Large Scale Assessment; *Michigan

ABSTRACT

Michigan school districts have been developing processes for supporting portfolios of student high school achievement since this was first mandated by law in 1991. The basic portfolio requirements have included transcripts, test scores, evidence of career planning, and anything the student added to show achievement. In the spring of 1995, Michigan collected 1,050 portfolios from eleventh graders in 42 schools as a 1% sample of the first student cohort expected to have and use portfolios. A report of findings, here summarized, was released in winter 1995. A portfolio scoring system had been developed in association with the Center for Research on Evaluation, Standards, and Student Testing at the University of California at Los Angeles. The value and difficulty of this Michigan model cannot yet be evaluated fully, but the portfolios that were submitted demonstrated that the school culture was not yet supporting the portfolios as anticipated. Professional development was needed to increase teacher understanding of the portfolio process and scoring. However, these first results showed the real feasibility of using hybrid review procedures and integrating the developed model. The answers to the skills demonstrated by Michigan students and whether the portfolio profiles can be considered true scores are not yet clear, but the potential for clarifying these issues is shown. (Contains three references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 102

What Have We Learned from Assessing Employability Skills Portfolios?

Catherine Smith
Michigan Educational Assessment Program
National Council on Measurement in Education
New York City, NY
April 10, 1996

Michigan school districts have been developing processes for supporting unique portfolios of student achievement across the 8th through 12th grades since this was first required in law in 1991. The basic portfolio requirements included transcripts, test scores, evidence of career planning and anything the student added to show achievement. This basic law was integrated with state efforts to help students in 8th through 12th grade discover, develop and document their employability skills--academic, personal management and teamwork skills--in unique student designed portfolios (Smith, 1993). All schools were required to support a portfolio process and most schools used employability skills assessment materials as part of their resource base. But did schools really help students self-assess employability skills as the schools met their own basic legal responsibility for a portfolio process?

In spring of 1995, the state of Michigan collected 1,050 portfolios from 11th graders in 42 Michigan schools, in a 1% sample of 11th graders in the first cohort expected to have and use portfolios. Portfolios were scored by teachers, counselors and other educators in summer of 1995. A summary report of findings was shared across the state in the winter.

What Did Michigan Hope to Learn from the Sample?

The Michigan Employability Skills portfolio assessment evidence model was developed to provide an opportunity for students to do meaningful self-assessment using the benchmarks and standards of evidence provided by the employment community and sound educational assessment practice. At the same time, the original task force that had recommended this assessment strategy had anticipated that schools would aggregate student profiles and use these as a basis for school improvement and that the state would create an indicator system based upon the evidence in student portfolios, to show Michigan's progress in developing a skilled workforce. This set of purposes led to several questions for the study of sampled portfolios to address:

- Can an indicator system be derived from a portfolio strategy based on student self-assessment and school improvement needs?
- Can two distinct scoring systems, one oriented to the student's needs and school improvement and the other focused on simpler indicator scoring, be combined in a coherent approach to train raters effectively?
- Do Michigan students show the profile of skills Michigan employers find critical?
- How will we know if we can trust the portfolio profiles as true scores of student skills?

Portfolio Scoring Systems: UCLA/CRESST and Michigan Model

The twin needs of student self-assessment and indicator information led state staff to work with staff at UCLA/CRESST on the type of scoring system needed for indicators of credible exhibits in student portfolios. The scoring system developed through UCLA/CRESST shared several premises with the Michigan evidence model, as described in Troper and Smith (NCME, 1995). These common premises made it feasible to consider a hybrid scoring strategy with the state sample.

TM025284

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
CATHERINE SMITH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Portfolio scoring had two parts: first an inventory of exhibits and a score of credible or not for each of the 14 skill areas Michigan and UCLA/CRESST staff agreed were expected in the portfolios (Troper et al, 1994). Then, for those portfolios with a credible exhibit and a written student analysis of the exhibit, an analytic scoring continued for the specific skills within the broad 14 areas using Michigan's Employability Skills evidence model.

The Michigan model had already been tried out by fifteen lead school districts using MDE procedures only in a pilot experience (MDE, Final Report, 1994). Data from that loose district level rating experience showed that the analysis required by the MDE model is valuable to raters in deciding scores but also that many schools have very primitive portfolios, often lacking analyses, and with exhibits clustered in only a few skill areas. Strong school effects in the portfolio processes and in the exhibits students were encouraged to include are evident, suggesting that the portfolio review may tell more about school effort than student skill at this early point. We expected, for this first cohort of portfolios, that the range of exhibits and analyses would be so narrow that most would not prove any skills using the demanding Michigan model alone. And yet, schools had done some baseline work in developing portfolio processes and providing basic exhibits for students to use, which should be captured as evidence of early effort. For these reasons and in order to build a state indicator system for Employability Skills, the CRESST model (which had been tried out and found reliable in a study of Los Angeles area teachers, Troper et al, 1994) was used in the 1995 Michigan sample as the first stage in a 2-stage rating procedure.

In stage 1, the UCLA/CRESST approach asks raters to examine each exhibit for its credibility for each of the 14 areas. Raters might find an exhibit credible for all, some, one or none of the 14 skill areas, using guidelines and some decision points defined in the training materials. For instance, any piece of writing other than poetry that is comprehensible to a rater is considered a credible direct demonstration exhibit of writing, whether it be a detailed report for a history class or a brief letter to a relative. A score at the satisfactory level or the 50th percentile in a recognizable mathematics test (whether it be an ACT score, PSAT, a state test score, etc.) is considered a credible third party exhibit in the broad area of mathematics. The number of credible exhibits is not an issue in the broad profile, just the presence of at least one credible exhibit. The profile of skills indicated by one or more credible exhibits is the student's final "score" from the UCLA/CRESST model. Each profile becomes an element in the aggregation of profiles across a state or other unit of concern for policy purposes.

In stage 2 of the hybrid scoring procedures, after assessing the credibility of exhibits alone, raters continue looking at the entries in the portfolio. Integrating the Michigan evidence model, raters determine if there is a written analysis making the student's case for including an exhibit as evidence of a particular benchmark in Michigan's Employability Skills. If an analysis is found, raters turn to the second page of the form, highlight the relevant benchmark area that the analysis suggested was the focus of the student's case, and look at the benchmarks to see if the student gets credit for any element in the skill area, putting the credible exhibit and the analysis statement together and weighing their evidence against the benchmark. In an analytic scoring of the evidence, a student might be credited with all, some, one or none of the specific skills s/he was hoping to prove. Finally, after reviewing the entire portfolio, raters write some advice to the student concerning at least two specific exhibits and concerning the overall portfolio's usefulness as evidence of skills. A student could have a generally credible exhibit, write an analysis that is not clear or detailed enough to show the specific relevance or quality of that exhibit, and score zero on the benchmark skills in that skill area. No cases were identified in the 1995 sample where a student received benchmark credit but did not meet the UCLA/CRESST standard for a credible exhibit; however, the training and rating procedures made that an unlikely event.

Issues in Rating Unique Portfolios

In this first year of review, portfolios were restricted in range, with very similar exhibits reflecting minimal requirements of law. They did not represent the full array of exhibits expected once school portfolio culture is more fully developed. Staff found that no portfolio in the sample was as complex or developed as many portfolios that had seen in the pilot review year, which was itself considered a modest rather than a full range of portfolios. Thus, the issue of restricted range is also an issue of school effects. Generally, the limited range and great similarity of the sampled portfolios suggest that these are not now valid indicators of student skill but rather that students are kept to a ceiling imposed by school procedures and understandings. Skills that cannot be proven with transcripts, test scores and school records are not likely to be proven by most students in this sample. Since the portfolios were not complex and because notching up teacher understanding to enhance culture was needed, planned review procedures were modified. Rather than employing independent review by raters acting alone after qualifying trials, each sampled portfolio was scored by two raters, mostly teachers or counselors, working together to agreement. Raters were debriefed both informally and with a response form about their understandings and rating problems.

Since the sample was drawn for indicator purposes at the state level, portfolio copies and scores were not returned to students and school scores were not created. Only state level indicators of Employability Skills were developed. In addition, the written advice to the student from the raters is being considered systematically at the state level to develop the activities and materials teachers need to mentor the portfolio process more effectively. The state scoring process is intended to serve as a model for districts attempting to rate portfolios for district or school information needed for school improvement and accreditation. It is not intended to develop numerical scores for students or for high stakes use regarding students or schools.

Profile of Skills Seen

Which skills did students demonstrate--and which were relatively neglected? The UCLA/CRESST approach indicates which exhibits the reviewers found at least minimally credible for a skill in the profile.

The UCLA/CRESST model was especially useful in producing indicators for state policy in this first year of portfolio review because the portfolios were not clearly organized or labeled for relevant Employability skill. The model allowed reviewers to acknowledge some effort by students without granting credit at the Michigan benchmark level. One valuable contribution of this indicator approach is the pattern shown: the indicator profile shows the skewedness of portfolio exhibits in this first cohort. Portfolios showed many more credible exhibits in Career Development, Writing, Reading, Math and Science and Technology than in the personal management and teamwork skills or the academic area of Problem Solving. (see Table 1)

This priofile ameliorates some concerns about Michigan's portfolio law, which expires July 1, 1996. That law does not appear to have stimulated development of portfolios across all the employability skills. Instead, schools made sure that student portfolios included minimal exhibits of academic areas (transcripts and test scores) and career development (plans and interest inventories) since these were explicitly required by the law. Although the assessment program has always encouraged schools to use the portfolio opportunity to help students document the full range of Employability Skills (and other desired outcomes as schools wish), the sample suggests this opportunity was not taken for the first cohort of students expected to have portfolios. Now, with the law expiring, schools can focus attention beyond the minimum requirements of law, into the more general school

improvement requirements. (Note: all schools are expected by Code to use authentic assessment to measure the readiness of students to assume adult roles and the strong School to Work message is that students show their understandings of work needs and their dawning skills over many years of experience.)

Table 1: Skills Shown by Credible Exhibits, By Portfolio

Any Exhibit Credible for Skill:	No	Yes
Writing	24%	76%
Speaking	85%	15%
Reading	45%	55%
Mathematics	43%	57%
Science/Technology	45%	55%
Problem Solving	93%	7%
Responsibility	66%	34%
Organization	68%	32%
Flexibility/Initiative	84%	16%
Career Development	23%	77%
Team Communicating	95%	5%
Responsiveness	94%	6%
Contributing	69%	31%
Membership	86%	14%

n= 1050 portfolios

The value and the difficulty of using the Michigan evidence model must also be also considered. The value is significant for those portfolios that included written analysis. Raters initially expressed concern with the UCLA/CRESST model and its seismograph approach to detecting credible exhibits, until they understood that these scores would only serve policy purposes and would not be returned to students. Raters did not want a student to conclude that a B in a general math class was sufficient math evidence for work readiness--and yet that was one credible exhibit. The Michigan model, with its analytic scoring at the level of individual skills constituting a benchmark, is a higher standard for students to reach and one that requires more complex exhibits and analysis. However, it is also difficult to see the connection students make between a complex exhibit (s) and a set of benchmarks skills unless the analysis is clear and detailed. Very few portfolios had such analyses. Most of the few first year analyses were neither clear nor detailed. The value and the difficulty of the Michigan model cannot be fully detailed until a broad range of portfolios is available for review; the current limited range does not allow reviewers to experience the anticipated issues with unique exhibits that led to development of the model.

What Did We Learn?

We established as the portfolios came in from schools, in uniform bundles of 20 to 30 that all looked alike by school, that schools were not yet supporting unique exhibits, student decisions about content or student written analyses of exhibits. While the assessment program had held countless workshops and provided extensive systematic materials statewide, school culture was not yet supporting the portfolios anticipated. Professional development was needed and the scoring of these relatively routine and minimal portfolios offered an opportunity to let school staff compare their products to the models envisioned and the standard expected.

Due to the strong emphasis on professional development for raters in the scoring, scoring was done by two person teams working to consensus and trading partners every few

portfolios. This social situation optimized the learning potential for each new scorer and ensured that scorers could hear different points of view about portfolios that seemed very limited in variety and credibility. The planned insertion of common portfolios, scored by all raters, did not produce enough scores for strong analysis of rater agreement, since raters ran out of time before enough common portfolios were scored. In fact, many of the "basic" sampled portfolios with only a few exhibits and no analysis had to be scored by the contractor after the official scoring sessions.

Only tentative comments are appropriate regarding the feasibility of scoring of unique exhibits. Most of these stem from one complex portfolio, not part of the sample, which was used as a common portfolio for scoring. This portfolio was remarked on by most raters, individually to staff and collectively in a last day session, as head and shoulders above the other portfolios. And yet, this level of skill and portfolio maturity is possible for many, if not most, students who have sufficient mentoring in portfolios. The student did not achieve all the benchmarks but he did have a credible exhibit in each of the 14 areas. He did have more than 20 analyses, which varied from clear and detailed to general and unhelpful. After rating this portfolio, most scorers said, "Now I know what I should be doing with my students!" They expressed a belief that most students could write at least the short and helpful analyses this student provided, with varying success. Raters remarked on the value added by analysis, both to the student's understanding and the scorer's understanding of context and purpose of the exhibit. The teacher who mentored that portfolio participated in the training and review and noted publicly that his was a good portfolio, but not unusual in her class of largely low SES rural students.

Raters faced some difficulties in consistently rating unique exhibits and wrote comments to the students about the need for organizing better and using analysis to make the student's purpose clearer. The written comments overall showed that raters understood the value a written analysis added to clarity and purpose of exhibits, even though they saw very few analyses and few truly unique exhibits in their scoring. The student analysis is at a very early stage of understanding in most schools in the state. This is clear from the statistic that only **337 exhibits, from 24 portfolios**, were accompanied by an analysis making a case for the student's skill, out of the 1050 portfolios scored. Further, many of these analyses were not, even in combination with a generally credible exhibit, sufficient for the student to earn any credit toward a benchmark skill. Students who wrote analyses were also concentrated in just two schools out of the 42 studied. (Note that none of the 42 was a pilot school from the benchmark or evidence review stages of the state assessment project, with the experience and portfolio culture that might have come from such pilot work).

Asking the scorers to write at least two specific comments on exhibits and one general comment on each portfolio slowed down the scoring and would be unnecessary if indicator development were the only focus of the study. Yet it had advantages as well. Writing comments gave every scorer practice in the kind of feedback to use if they engage in local review of portfolios or if they mentor portfolios. It allowed the assessment office to gauge the importance scorers attached to analyses as vital elements in portfolios. And it gave scorers a sense of openness, after being constrained to use UCLA's system and MDE's system of what is credible and convincing--they liked being able to express their uncertainties directly.

Aggregating Portfolio Results for State Indicators

The possibility of aggregating findings from unique student portfolios to add to system (state or district) knowledge about its success in preparing students for the workforce is full of unknowns. As districts and states attempt to examine their students' preparation for the next stage of school or life (not just for employment), such assessment and aggregation procedures are likely to be proposed, for regional School to Work efforts and related policy

needs. The employability skills portfolio experience offers some preliminary data on the viability and the problems of such procedures. One advantage to the use of aggregated state data, apart from the cost advantages compared to more intensive sampling, is that the point can be made that all students in a state need to be made work ready and that all kinds of students are included in the profile. Table 2 shows the demographic profile of the first state sample. Inclusion of curriculum focus of students helps to get the point about inclusiveness across in policy reporting. Yet the state sample is too small to break out distinctive profiles by race or curriculum; such break out could be feasible for gender, if a policy purpose would be served. For this first year, with strong school effects and low range in the portfolios, break outs even of a larger sample would make little sense.

Table 2: Demographics of Sample

Student Characteristics	Number of Students
GENDER:	
Male	465
Female	562
Not Identified	23
CURRICULUM:	
Special Education	29
Career/Vocational	165
General Ed Only	853
Not Identified	12
RACE/ ETHNICITY:	
White	845
Black	91
Hispanic	18
Asian	12
American Indian	11
Not Identified	73

Conclusion:

The first review of a state sample of unique portfolios for evidence of Employability Skills showed the real feasibility of using hybrid review procedures, integrating UCLA/CRESST's seismograph approach to the fourteen broad areas with the Michigan evidence model's demands for analysis and potential for analytic scoring of portfolios including both exhibits and analyses. Portfolios planned for self--assessment can be skimmed for indicators of interest at a policy level. Raters saw the consistency of expectations across the indicator procedures and the Michigan evidence model procedures and did not express, either in training or in writing, reservations about the compatibility of the scoring systems. However, the statistical agreement of raters on scoring decisions cannot be estimated from the first year study due to the decision to emphasize professional development in the rating sessions, which was stimulated by the narrow and low range of portfolios collected in 1995. Further work on agreement will be possible as the range of evidence broadens and the scoring model shifts from professional development needs.

The answers to the other two questions asked, the extent of skills shown by Micihigan students and the issue of whether these profiles can be considered true scores, remain on the table. Because of the narrow range of exhibits and the few analyses, it is premature to decide that Michigan students lack employability skills. Future samples with a broader range of school implementation models and with clearer evidence of student variation within schools is needed before conclusions about student skills are made. The potential for using the portfolio as a basis for self-assessment, school improvement or state indicators rests largely with school decisions about implementation and with school culture about

portfolio work. If schools do not support students in using and learning about their strengths through portfolio self-assessment, these folders will not be representative of students actual skills.

Appendix: Reviewers Reactions to Portfolios
These were drawn at random from a stack of handwritten reviewer comments by the teacher-reviewers in the summer review of the 1% sample of 11th grade portfolios.

"Too much emphasis on testing with no follow through in the form of reflective analysis."

"You have some excellent employability and career planning exhibits and your test scores will help you plan your future. You might expand the scope of your portfolio by adding some other kinds of exhibits which give a reviewer a more complete picture."

"No analysis was given for any exhibit. You need to draw conclusions for us. Why did you include these? What does this tell about you?"

"You have included some worthwhile exhibits but too many of them are repetitions, for instance, the report cards which reflect the same grades. Try to include a greater variety of skills and direct exhibits."

"Your cross-age tutoring activities seem to be very important to you. You have surely demonstrated some skills that would impress a prospective employer. Creating an analysis of what you did with your young charges will show the specific skills that you have."

"Records and certificates are great, but you need to show what you did to win these awards. Please include an analysis with each award because there is not enough to show representation of any of your own work."

"This is a very strong portfolio. The writing samples, letter of recommendation and Australia/ New Zealand trip all are especially strong exhibits. "

"This is a good beginning to a portfolio. You should consider including more samples of academic work such as math and science... elaborate on how your team participation has shown teamwork skills."

Sources:

Michigan Department of Education. Final Report: Portfolio Project. September, 1994.

Smith, Catherine. "Assessing Job Readiness Through Portfolios," The School Administrator. December, 1993. 26-31.

Troper, Jonathan, Eva L. Baker and Catherine B. Smith, "Using the Employability Skills Portfolio Inventory: Training Instructions for Trainers and Portfolio Reviewers." UCLA, CRESST. Working paper. July, 1994.